



Développement et validation psycholinguistique d'un corpus pour l'étude des bases neurales de la prosodie affective

Virginie Beaucousin, Anne Lacheret-Dujour, Marie-Renée Turbelin, Michel Morel, Bernard Mazoyer, Nathalie Tzourio-Mazoyer

► To cite this version:

Virginie Beaucousin, Anne Lacheret-Dujour, Marie-Renée Turbelin, Michel Morel, Bernard Mazoyer, et al.. Développement et validation psycholinguistique d'un corpus pour l'étude des bases neurales de la prosodie affective. 25èmes Journées d'Etudes sur la Parole, Apr 2004, Fès, Maroc. [4 p.]. hal-00103659

HAL Id: hal-00103659

<https://hal.science/hal-00103659>

Submitted on 4 Oct 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Développement et validation psycholinguistique d'un corpus pour l'étude des bases neurales de la prosodie affective

*Beaucousin V. *, Lacheret A. **, Turbelin MR. *, Morel M. **, Mazoyer B. *, Tzourio-Mazoyer N. **

CRISCO**

Université de Caen, Caen 14032 Cedex, France

GIN, UMR6194*

GIP Cyceron, blvd H. Becquerel, BP5229, 14074 Caen cedex, France

Tél.: ++33 (0)2 31 56 56 27 - Fax : ++33 (0)2 31 56 54 27**

Tél. : ++33(0)2 31 47 02 60, Fax : ++33 (0)2 31 47 02 22*

Mél: lacheret@crisco.unicaen.fr - <http://www.crisco.unicaen.fr>

Mél : beaucousin@cyceron.fr, <http://www.cyceron.fr/gin/>

ABSTRACT

This study is carried out in the framework of research on affective prosody. It concentrates on the role of prosody in the recognition of various types of expressive messages. The work presented here, which is the fruit of the collaboration between cognitive neuroimaging and linguistic research teams located at the University of Caen, is aimed at elaborating and validating linguistic material for the study of the neural bases of emotions and attitudes with functional magnetic resonance imaging.

1. INTRODUCTION

Un champ de recherche très large est ouvert aujourd'hui à l'interface de la linguistique, de la phonétique et des neurosciences qui a pour but de comprendre comment la prosodie, au même titre que la syntaxe et la sémantique, participe activement à la construction du discours et à son interprétation. Dans ce contexte, le travail présenté ici s'inscrit dans le cadre d'une collaboration entre deux laboratoires de recherche de l'université de Caen, le GIN, laboratoire de neuroimagerie, et le CRISCO, centre de recherche inter-langues sur la signification en contexte. Faisant dialoguer neuroimagerie, psychologie, traitement du signal, phonétique et linguistique, cette entreprise pluridisciplinaire a pour objet les bases neurales des prosodies affectives. Ces dernières interagissent à la fois avec les fonctions linguistiques traitées par l'hémisphère gauche, et la pragmatique qui recruterait des aires de l'hémisphère droit. Une telle étude devrait ainsi permettre d'appréhender la coopération hémisphérique mise en jeu au cours du traitement du discours. De plus, en neurosciences : le réseau neural de la prosodie émotionnelle est mal connu et la prosodie attitudinale n'a pas été étudiée à notre connaissance (pour une revue voir Beaucousin & al., [4]). En pratique : qu'en est-il de la perception et de l'identification des émotions et des attitudes dans l'activité de langage ? De quelle manière les informations prosodiques (fréquence fondamentale, intensité, durée et timbre) contribuent-elles de façon

significative à ce processus d'identification perceptive ? Au-delà de cet ensemble de questions, que peut-on dire sur le caractère autonome ou intégré du traitement prosodique par rapport aux autres niveaux de traitement dans la perception du discours expressif ? Le but de notre communication est d'exposer la première phase de ce travail : la mise au point du matériel expérimental utilisé pour aborder ces différentes thématiques et sa validation psycholinguistique. Nous présentons en premier lieu les caractéristiques du corpus sous l'angle pragmatique et verbal : types expressifs retenus pour l'analyse et prosodie associée, choix du matériel lexical et des constructions syntaxiques, méthode utilisée pour la production et l'enregistrement des stimuli (parole synthétique vs voix d'acteurs). Dans un second temps, nous donnons les résultats des analyses perceptives conduites sur 16 sujets pour le contrôle et la validation du matériel linguistique.

2. PRÉSENTATION DU MATÉRIEL LINGUISTIQUE

Notre objectif général étant une meilleure compréhension de la prosodie affective et de ses différentes composantes, il s'agit de confronter des conditions avec ou sans prosodie pour évaluer l'apport de la prosodie affective dans le discours : Beaucousin [3]. Nous présentons ici les critères retenus pour la construction de notre corpus de 320 phrases : 120 stimuli émotionnels, 120 attitudinaux, 80 neutres.

2.1. Quelles attitudes, quelles émotions ? Regard pragmatique et prosodique

Dans l'état actuel des connaissances sur la parole expressive, il est communément admis que celle-ci véhicule deux types d'affects : les émotions et les attitudes. Les premières, qui servent d'indices à l'auditeur pour évaluer l'état psychologique du sujet parlant, sont véhiculées par différents canaux verbaux et co-verbaux (gestes, mimiques faciales, tremblements, modes articulatoires, prosodie, etc). Les secondes relèvent de **signaux** délibérés. Mais pour être identifiés comme tels,

encore faut-il que ces signaux répondent à un certain nombre de schémas conventionnels, *i.e.* partagés par une communauté.

Malgré la diversité des travaux et des modèles proposés (Aubergé [2], Morel & Bänziger [8]), force est de constater d'abord qu'il n'est pas évident de poser une frontière nette entre émotions et attitudes (Lacheret & Beaugendre [6]). En supposant que cette frontière soit établie, un second problème se pose : il n'existe pas de consensus ni sur le type ni sur le nombre d'attitudes et d'émotions pouvant clairement être distinguées dans la perception du langage. Et pour cause, les critères et les méthodes utilisés pour faire émerger différents types émotifs et attitudeaux sont extrêmement instables, dépendant intimement de l'angle d'attaque retenu pour l'analyse. Bref, dans ce domaine foisonnant, on est loin d'avoir atteint un minimum épistémologique commun, utilisable pour quiconque entreprend une recherche sur la parole expressive. Fors de ce constat, nous avons opté pour la démarche suivante : (i) partir des classifications qui se retrouvent de manière relativement stable d'un modèle à un autre, (ii) restreindre notre paradigme par l'utilisation d'indices prosodiques tels qu'ils sont posés dans la littérature de façon à limiter le plus possible la confusion des genres dans une tâche de reconnaissance en IRMf. En pratique, si on retrouve de manière constante dans les travaux sur les émotions, 6 types : la colère, la tristesse, la joie, la peur, le dégoût et la surprise (Diaferia [5]), nous avons sélectionné trois émotions primaires : gaieté, tristesse et colère sur les bases de contrastes mélodiques a priori bien identifiés et distincts entre eux, et s'écartant également du patron intonatif porté par une phrase assertive neutre. Concernant les attitudes, nous avons retenu le doute, l'évidence et l'ironie, en nous fondant sur les expériences réalisées à l'ICP de Grenoble qui attestent de l'identification précoce de ces trois types dans des tâches de gating : Aubergé & al. [1]. Soit 120 phrases pour les émotions (40 stimuli pour chaque type émotionnel), 120 phrases pour les attitudes (40 stimuli pour chaque type attitudeal) et 80 phrases de référence dites « neutres ».

2.2. Composantes syntaxique et lexicale

Lors du choix des constructions syntaxiques et du matériel lexical, il s'est agi pour nous de trouver le compromis le plus judicieux, face à deux contraintes majeures : (i) contraintes écologiques : énoncés aussi naturels et variés que possible, (ii) contraintes du paradigme d'IRMf : équivalence temporelle des stimuli, patrons syntaxiques équivalents pour chaque catégorie, même nombre de stimuli pour chaque catégorie.

Les constructions syntaxiques s'articulent autour de trois types quantitativement homogènes :

- type canonique SVO (CIR)
- structures détachées à gauche
- structures détachées à droite

Table 1 : Exemples d'énoncés pour les 3 émotions.

Gaieté	tristesse	colère
<i>Le copain d'ma fille, il est génial</i>	<i>Les fins de mois, elles sont dures, très dures</i>	<i>Ce chauffard, il m'a renversé et ne s'est même pas arrêté</i>

Table 2 : Exemples d'énoncés pour les 3 attitudes.

Doute	évidence	ironie
<i>Lui, il a entendu un chien parler</i>	<i>Un oiseau, ça vole</i>	<i>Ce feignant, il a travaillé avec zèle</i>

Sous l'angle lexical, nous devons sélectionner un matériel, non ambigu, *i.e.* a priori compatible avec un et un seul type de prosodie affective.

2.3. Enregistrements des stimuli : voix synthétique vs. naturelle

Afin de disposer d'une condition de référence, nous avons utilisé le système synthèse de parole à partir du texte Kali (Morel & Lacheret [7]) qui, comme toutes les synthèses de ce type à l'heure actuelle, reste extrêmement limité sous l'angle communicationnel : doté d'une prosodie générée sur les bases de contraintes syntaxiques et rythmiques, il ne dispose d'aucune balise pragmatique. La moitié des 320 phrases ont donc été lues par une voix synthétique masculine, le reste étant prononcé en utilisant une des voix féminines de notre synthétiseur. Les mêmes phrases ont été enregistrées dans une chambre sourde par 2 acteurs (un homme, une femme).

3. VALIDATION PSYCHOLINGUISTIQUE

De manière à valider le corpus de phrases, nous avons sollicité 16 sujets (9 hommes et 7 femmes, droitiers, âgés de 27 ± 5 ans (moyenne \pm écart-type)).

3.1. Acquisition des données

L'acquisition des données a eu lieu au cours de deux sessions différentes d'environ 45 minutes chacune, en commençant par les phrases produites par Kali pour finir par les stimuli actés. Les sujets devaient classer les phrases entendues dans les différents types, toute catégorie confondue (gaieté, colère, tristesse, doute, ironie, évidence, neutre), par l'intermédiaire d'une réponse sur le clavier de l'ordinateur. Les phrases étaient présentées en stéréo, elles duraient en moyenne 3 secondes et les sujets n'avaient pas de temps limité pour la réponse qui, comme les temps de réactions, a été enregistrée. La présentation des phrases et l'enregistrement des réponses ont été réalisés avec le logiciel SuperLabtm Pro version 2.0 (CEDRUS, <http://www.superlab.com/papers/>). Un questionnaire post-expérimental a permis de recueillir les impressions

des sujets sur la compréhension des phrases et la difficulté de la tâche à la fin des sessions de test.

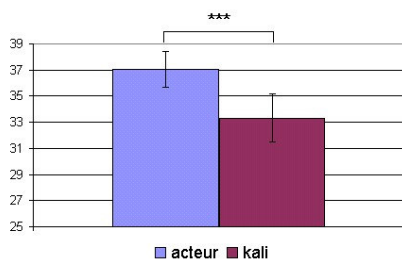
Pour l'analyse des données, nous avons analysé la variable correspondant au nombre de phrases correctement classées (résultat sur 40 phrases au total pour chaque catégorie). Cette variable, ne suivant pas une loi normale, a été analysée avec une statistique non paramétrique (test de rang de Wilcoxon). Le temps de réaction a également fait l'objet d'une analyse statistique paramétrique de type ANOVA à mesures répétées puisque suivant une loi normale. Pour cela, deux sujets ont été écartés car ils augmentaient la dispersion de la variable qui ne suivait plus une loi normale. L'analyse portait sur le facteur voix (phrases lues par les acteurs versus KALI).

3.2. Résultats

Nous voulons tout d'abord préciser que, lors de l'examen du questionnaire post-expérimental, il est apparu que tous les sujets avaient eu des difficultés à distinguer la modalité neutre de la modalité d'évidence. En conséquence, si nous avons conservé ces données dans les analyses globales, nous ne donnerons pas ici les résultats spécifiques de ces deux catégories, trop indiscernables l'une de l'autre. Finalement, les différentes catégories émotionnelles traitées sont la colère, la gaieté, et la tristesse, les catégories attitudinales : le doute et l'ironie.

Comparaison du nombre de réponses correctes en fonction de la présence de prosodie affective

Si globalement les performances des sujets sont correctes, une différence significative a été détectée en fonction du type de voix, le nombre de phrases bien classées étant plus élevé lorsqu'elles sont pourvues d'une prosodie affective (nombre de bonnes réponses pour les phrases dites par les acteurs $37,1 \pm 1,38$ (Moyenne \pm SEM) que pour les phrases synthétiques $33,3 \pm 1,84$; $p < 0.001$; voir graphique 1).

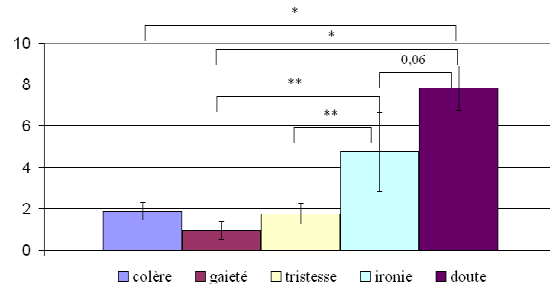


Graphique 1 : Moyenne \pm SEM du nombre de bonnes réponses sur 40 pour les phrases lues par les acteurs (en bleu), et celles produites par KALI (en bordeaux) toutes catégories confondues (***, $p < 0.001$).

Rôle de la prosodie affective pour le classement des différentes catégories

Le calcul de la différence entre le nombre de bonnes réponses de classement sur les phrases avec prosodie

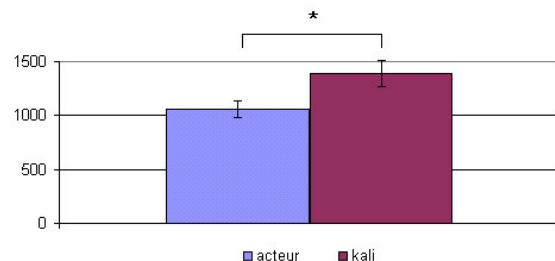
affective (dites par les acteurs) et les phrases ne contenant que la prosodie linguistique (phrases générées par le logiciel KALI) permet de mieux comprendre pour quelles catégories la présence de la prosodie améliore le classement. L'effet porte principalement sur les attitudes : doute et ironie (voir graphique 2).



Graphique 2 : Moyenne \pm SEM de la différence du nombre de bonnes réponses entre les phrases lues par les acteurs, et celles produites par KALI® pour chaque catégorie (Test de Wilcoxon, * $p < 0.05$; ** $p < 0.01$).

Etude de temps de réaction en fonction de la présence de prosodie affective

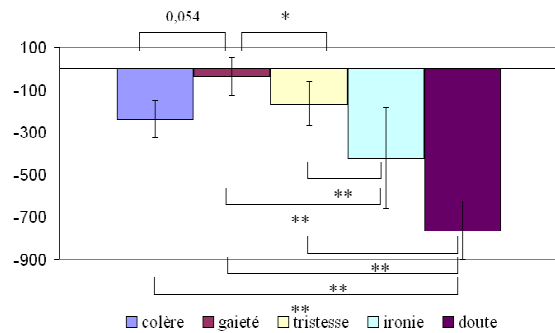
Cette comparaison a été effectuée sur les phrases correctement classées. Elle met en évidence un effet de la prosodie affective : les phrases énoncées par les acteurs sont classées plus rapidement que les phrases produites par le logiciel KALI (temps de réponse pour les phrases dites par les acteurs 1059 ± 82 ms (Moyenne \pm SEM) ; pour les phrases produites par KALI 1389 ± 126 ; $p < 0.02$; graphique 3).



Graphique 3 : Moyenne \pm SEM du temps de réaction en millisecondes pour les phrases correctement reconnues lues par les acteurs (en bleu), et celles produites par KALI (en bordeaux) toutes catégories confondues (*, $p < 0.05$).

Sur l'ensemble des phrases bien classées, le calcul de la différence en terme de temps de réaction entre le classement des phrases avec prosodie affective et celles qui en sont dépourvues, correspondant au coût cognitif de l'absence de prosodie affective, nous a permis de mettre en évidence une augmentation importante du temps de réaction pour le classement des phrases attitudinales (entre 400 et 700 ms supplémentaires). Le classement des phrases à contenu émotionnel est lui aussi plus lent pour la colère et la tristesse, mais la différence est limitée à 200 ms. On note néanmoins qu'ironie et colère ne présentent pas de différence en terme de temps sup-

plémentaire de traitement, et qu'au sein des émotions il existe des différences significatives.



*Graphique 4 : Moyenne \pm SEM de la différence du temps de réaction en millisecondes entre les phrases correctement reconnues avec prosodie affective, et celles sans prosodie affective pour chaque catégorie (Test de t apparié, * $p < 0.05$; ** $p < 0.01$).*

3.3. Discussion

1. L'utilisation de la synthèse de la parole dotée d'une prosodie a minima (dont la génération ne repose que sur des règles syntaxiques et rythmiques) nous a permis de confirmer le rôle significatif de la prosodie dans l'identification des émotions et des attitudes. Celle-ci permet à la fois de mieux classer les différentes catégories, et donc de restreindre l'ambiguïté du discours mais également, pour les phrases bien classées, de diminuer le coût cognitif du traitement de la partie affective du discours. Ceci étant vrai aussi bien pour les émotions que pour les attitudes.

2. Le moins bon taux de reconnaissance des attitudes et la nécessité d'un temps de traitement plus long lorsque les phrases sont dépourvues de prosodie affective démontre l'importance de la prosodie pour une compréhension correcte des énoncés produits. Si les émotions peuvent être raisonnablement comprises, même en l'absence d'une prosodie adaptée, le nombre d'erreurs au cours de la perception des attitudes augmente et une bonne interprétation se fait au prix d'un effort cognitif important. Ceci est à la fois illustré par les résultats présentés ici et par les réponses des sujets aux questionnaires post-expérimentaux qui signalent tous une difficulté accrue pour le traitement des phrases sans prosodie affective et particulièrement pour les attitudes. Néanmoins, l'étude des temps de réaction sur les réponses correctes permet de nuancer la distinction entre émotion et attitude : il existe des différences au sein même des émotions, et il n'y a pas de différence entre la colère et l'ironie. Cependant, ces résultats devraient être confirmés par une étude supplémentaire palliant au problème de l'ambiguïté évidence / neutre qui pourrait avoir influencé plus particulièrement les temps de réactions.

3. Ces résultats indiquent que la distinction opérée par la linguistique entre attitudes et émotions se traduit, dans une certaine mesure, par une différence de traite-

ment sur le plan de la psychologie expérimentale, ce qui justifie l'étude des bases neurales du traitement des émotions et des attitudes. D'autre part, nos données démontrent que l'utilisation d'un outil de synthèse vocale est une approche pertinente pour évaluer les différences existantes entre ces deux types de prosodies affectives. Enfin, ce travail nous a permis de finaliser un sous corpus de 180 phrases homogènes pour préparer le paradigme d'imagerie fonctionnelle (nous avons éliminé les phrases mal entendues et les phrases mal classées). Les résultats de l'étude préliminaire en IRMf seront présentés à Human Brain Mapping 2004.

BIBLIOGRAPHIE

- [1] V. Aubergé, N. Audibert, A. Rilliard. Can we perceive attitudes before the end of sentences? A gating paradigm for prosodic contours. *EUROSPEECH*, Rhodes, 1997.
- [2] V. Aubergé (dir.). *Journée Parole expressive*. Université de Grenoble, novembre 2004.
- [3] V. Beaucousin. Les bases neurales de la compréhension des prosodies affectives. Mémoire de DEA de biologie cellulaire, Université de Caen, 2003.
- [4] V. Beaucousin V, A. Lacheret-Dujour, N. Tzourio-Mazoyer. La prosodie. In *Cerveau et langage*. Hermès, Paris, 223-247, 2003.
- [5] M.L. Diaferia. Les attitudes de l'anglais : vers des indices prosodiques. DEA de Sciences cognitives, INPG, Grenoble, 2001.
- [6] A. Lacheret et F. Beaugendre. *La prosodie du français*, Editions du CNRS, Paris, 1999.
- [7] M. Morel et A. Lacheret. Kali, synthèse vocale à partir du texte : de la conception à la mise en oeuvre. *Traitement automatique des langues, synthèse de la parole à partir du texte*, 42, Ch. D'Alessandro (éd.). Hermes, Paris, 193-221, 2001.
- [8] M. Morel et T. Bänziger. Le rôle de l'intonation dans la communication vocale des émotions. A paraître dans *CILL* 30 n°1-3, 2004.